Imprecisely georeferenced specimen data provide unique information on species' distributions and environmental tolerances: Don't let the perfect be the enemy of the good

Running title: Using imprecisely geolocated occurrences

Authors:

Adam B. Smith[1*] [ORCiD: 0000-0002-6420-1659]
Stephen J. Murphy[1,2] [ORCiD: 0000-0001-8679-2851]
David Henderson[3] [ORCiD: 0000-0003-3523-0175]
Kelley D. Erickson[1] [ORCiD: 0000-0001-7980-9985]

[1] Center for Conservation and Sustainable Development, Missouri Botanical Garden, 4344 Shaw Boulevard, Saint Louis MO 63110 USA

[2] Present address: The Ohio State University, Department of Evolution, Ecology, and Organismal Biology, 318 W 12th Avenue, Columbus, OH 43210 USA, murphy.1132@osu.edu

[3] Department of Biology, Washington University in Saint Louis, 1 Bookings Drive, Saint Louis MO 63130 USA, hdavid@wustl.edu

* Corresponding author

Abstract

Aim: Conservation assessments frequently use occurrence records to estimate species' geographic distributions and environmental tolerances. Typically, records with imprecise geolocality information are discarded before analysis because they cannot be matched confidently to environmental conditions. However, removing records can artificially truncate species' environmental and geographic distributions. Here we evaluate the trade-offs between using versus discarding imprecise records when estimating species' ranges and climatic tolerances.

Location: North America.

Time period: 1970-2019 and 2061-2080.

Methods: We collated records from 44 species in the genus *Asclepias* (milkweeds). Records were designated "precise" if they could be matched confidently to environmental data, and "imprecise" if not. We compared estimates of extent of occurrence (EOO), climatic niche breadth, and exposure to climate change using precise records only, as well as precise plus imprecise records together. To estimate EOO, we conservatively assigned imprecise records to points within their area of likely collection that were closest to the centroid of precise records. Similarly, to estimate climatic tolerances and exposure to climate change, we matched imprecise records to climate values that were most similar to the mean across precise records.

Results: Across all species, including imprecise records increased EOO by 85% (median value; range across species: 0-2011%). Univariate niche breadth in mean annual temperature and precipitation increased by 25% (0-353%) and 28% (0-292%), respectively, while multivariate niche volume increased by 175% (8-13909%). Adding imprecise records increased suitable area in the present and area that remained suitable in the future.

Main conclusions: Imprecise records provide novel information about species' distributions and climatic niche tolerances. While the default practice of discarding imprecise records ensures that only accurate data are used, it dramatically reduces estimates of range size and overestimates exposure to climate change. The benefits of discarding imprecisely geolocated records must be balanced against the loss of information incurred by their elimination.

Keywords [6-10, alphabetically, should repeat keywords in title and abstract]

climate change vulnerability, coordinate uncertainty, georeferencing, niche breadth, natural history museum specimen records, niche truncation, species distribution model

Introduction

Accurately identifying species' geographical distributions and environmental limits is fundamental to assessing the species' current conservation status and vulnerability to anticipated climate change (Foden & Young 2016; IUCN 2019). However, acquiring data that can reliably indicate where species occur across space and time remains highly challenging (Feeley & Silman 2010, 2011; Moudrý & Šímová 2012; Collins et al. 2017). Conservation assessments frequently rely on museum and herbarium specimen records to assess conservation status (Miller et al. 2012; Young et al. 2016; IUCN 2019) and exposure to expected climate change (Dawson et al. 2011; Pacifici et al. 2015; Foden & Young 2016). Using occurrence records from museum and herbaria to assess conservation status and climate change exposure requires two key assumptions that are often difficult to meet. First, records need to reflect environmental conditions experienced by the species at the place of observation (Graham et al. 2008). When records can only be geolocated to a general area (e.g., a protected area or geopolitical boundary), it becomes impossible to confidently assign a specific environmental datum to each record (Graham et al. 2008; Collins et al. 2017). Second, when the goal is to estimate environmental tolerances, records should encompass the full suite of conditions comprising the species' fundamental niche (Thuiller et al. 2004; Peterson et al. 2018). If available occurrences represent only a portion of the fundamental niche, then the species' actual environmental tolerances will be underestimated (Thuiller et al. 2004; Qiao et al. 2019).

Importantly, decisions regarding the spatial precision of occurrence points create a trade-off between utilizing a sufficient number of records to represent the full geographic and environmental distribution of a species, and ensuring a high-fidelity match between occurrences and environmental data. On the one hand, the use of imprecise records can inflate niche breadth

(Feeley & Silman 2010; Collins et al. 2017) and reduce the predictive accuracy and inferential power of ecological niche models (Graham et al. 2008; Fernandez et al. 2009; Tulowiecki et al. 2015; Osborne & Leitão 2009; Gábor et al. 2020; Mitchell et al. 2016; Cheng et al. 2021). These issues arise because environmental values to which imprecise records should be matched remains uncertain. Hence, geographic imprecision yields environmental imprecision, thereby "fuzzing" the observable relationship between occurrence and environment (Collins et al. 2017; Park & Davis 2017; Pender et al. 2019). Concurrently, niche breadth might also be inflated if retroactively georeferenced (and thus less precise) records are older and represent populations that have since been extirpated due to contemporary climate change. In this case, assuming the climates associated with these records represent suitable conditions will overestimate actual environmental tolerances (cf. Escribano et al. 2016; but also see Faurby & Araújo 2019).

In contrast, discarding geographically imprecise records can also pose substantial risks. Eliminating imprecise records can exacerbate undersampling of species locations and environments, even if the exact locations and conditions to which such records should be matched are unknown (Meyer et al. 2016). This is especially of concern if precisely-georeferenced records do not capture the full range of tolerable environments. When this is the case, niche models and other techniques for estimating environmental tolerances will underestimate actual tolerances, thereby inflating perceived exposure to climate change and other threats. These risks are especially great for rare species that tend to be represented by a small number of specimens (Lomba et al. 2010; Sheth et al. 2012; Zizka et al. 2018). Indeed, recent assessments of available data reveal that a substantial proportion of records in publicly available databases can only be located very generally (40-90%, depending on the source) due to missing or large coordinate uncertainty values (Collins et al. 2017; Moudrý & Devillers 2020; Marcer et

al. 2021). Concerns regarding the removal of records from biogeographic analysis are only rarely considered in previous studies. Indeed, in a survey of the literature, we found that half of all studies using museum or herbarium specimen data for species distribution modeling or ecological niche modeling opted to discard spatially imprecise records (Fig. 1; Appendix S1). The remainder of studies either retained imprecise records without using methods to account for uncertainty, or had no imprecise records to discard.

Given the importance of both data quantity and quality, we believe that a reexamination of the benefits and risks of removing imprecise specimen records before conducting conservation assessments is timely. Here we explore the trade-offs between retaining and discarding imprecise records using 44 species in the genus *Asclepias* (milkweeds; family Apocynaceae) native to North America. We classified records as "precise" if they had spatial uncertainty small enough to enable them to be matched with confidence to environmental data, and "imprecise" if they could not. We emphasize that our definition of an imprecise record does not include records with locations appearing to be outside the range of the species (i.e., geographic outliers; Feeley & Silman 2010) or specimens that do not pass quality-assurance checks (Chapman 2005). We then compared estimates of range size and climatic niche breadth of each species when using only precisely georeferenced records to estimates using precise and imprecise records together. We also compared potential climate change exposure (i.e., the magnitude of climate change experienced by a species; Williams et al. 2008; Dawson et al. 2011) by projecting ecological niche models calibrated using either precise records or precise plus imprecise records to future climate scenarios.

Methods

Specimen data and cleaning

We manually downloaded all available specimen data for *Asclepias* from the Global Biodiversity Initiative Facility (GBIF) for North America (Canada, the United States, and Mexico). Detailed procedures for cleaning and classifying the records are described in Appendix 2. Briefly, we first removed records missing species names and records based on human observations. We then discarded records which could not be unambiguously assigned a collection year or that were collected before 1970 (the first year represented in the climate data product utilized in this study). We also removed records that were likely cultivated, purchased, or otherwise collected in conditions not indicative of natural environmental requirements. Names of states/provinces and counties were manually cleaned then matched with names used in the Database of Global Administrative Areas (GADM ver 3.6; www.gadm.org; downloaded May 20, 2020). We removed records that could not be located to at least a state/province or occurred outside of coterminous North America and minor outlying islands. We then eliminated records containing conflicting information, such as those collected in a county that did not appear in the given state, or coordinates that did not fall within the given county or state.

Next, we assessed coordinate uncertainty for the remaining records. Coordinate uncertainty is typically reported as the radius of a circle with a center at the given coordinate pair (Wieczorek et al. 2004; Chapman & Wieczorek 2000), although not all records with coordinates report uncertainty in GBIF (Moudrý & Devillers 2020). Uncertainty reflects imprecision in the method of geolocation (e.g., GPS versus maps-based interpolation), imprecision in the locality description (e.g., “near Gettysburg” versus “7.3 miles northeast of Gettysburg on Old Harrisburg Road”), and other aspects related to coordinate reference systems and methods of collection (Wieczorek et al. 2004; Chapman & Wieczorek 2000). We categorized records into two groups,

"precise" and "imprecise", based on the confidence with which they could be matched to the resolution of our climate data. Precise records were those with an associated coordinate uncertainty of ≤5000 m, in accordance with guidelines used for calculating climate change vulnerability (Young et al. 2016; see also Graham et al. 2008). Imprecise records comprised records that had a coordinate uncertainty >5000 m, or that could only be located to a geopolitical unit (county/parish or state/province). Records that could not be assigned to either one of these categories were removed. We also removed records with coordinates that had an area of uncertainty larger than San Bernardino County (the largest "county"-level geopolitical unit in the coterminous US at 52,104.5 km^2) to exclude records which we considered to be too imprecise to convey salient information about the species' relationship to the environment. Maps of each species' records were then visually inspected, and geographic outliers were scrutinized and removed if appropriate (e.g., one georeferenced record was from a bridal bouquet, and several others were purchased or cultivated). We defined duplicate records in different ways depending on the type of records then kept just one of each set of duplicates (Appendix S2). Duplicate records of each species were discarded. Finally, we removed species for which we had fewer than 5 precise records, resulting in 44 total species for analysis.

Extent of occurrence

The geographic extent of occurrence (EOO) of a species is defined as the area of the minimum convex polygon circumscribing all occurrences. EOO is used as an index of the degree to which risks from threats are spread among populations (IUCN 2019). For each species, we compared the EOO estimated from precise records to those using precise plus imprecise records. To calculate the minimum convex polygon when imprecise records were included, we used the coordinate locations of precise records as they were reported, but for imprecise record we used

the point on the border of the polygon that represented its likely collection locality that was closest to the geographic centroid of the precise records (Fig. 2a). This yielded the smallest possible EOO for precise plus imprecise records. We removed areas covered by major water bodies (oceans and the Great Lakes) before calculating areas of EOOs. Differences in EOO calculated without or with imprecise records were evaluated with a paired Wilcoxon signed-rank test.

Environmental data

We use interpolated climate coverages from WORLDCLIM Ver. 2.1 (Fick & Hijmans 2017) at 10-arcmin resolution (~18.5 km latitudinally) for the periods 1970-2000 and 2061-2080 from CMIP6 under Representative Concentration Pathways (RCPs) 4.5 and 8.5 (achievable under Shared Socioeconomic Pathways 245 and 585; O'Neill et al. 2015). For future climate, we used the ensemble mean predicted climate across six earth system models (Appendix S3).

Univariate and multivariate niche breadth

We compared climatic niche breadth in mean annual temperature (MAT) and mean annual precipitation (MAP) using just precise records versus using precise plus imprecise records. Niche breadth was calculated as the range of values of MAT or MAP across all occurrence records for each species. We also calculated the multivariate climate niche volume and surface area of this volume from the convex hull circumscribing all occurrences in the first three axes of a principal components analysis (PCA) on all 19 BIOCLIM variables (Nix 1986). The PCA was calculated across all 10-arcmin cells in North America, and the first three axes explained 88% of the total variation. We compared differences in univariate and multivariate niche breadth between precise and precise plus imprecise records using paired Wilcoxon signed-rank tests.

Assigning climate data to imprecise records requires making a decision about how to "locate" a record in climate space from the climates in the area from which the record was collected. For the analyses presented in the main text, we chose climate values that were closest to the mean value estimated across all precise records. This method provides the most conservative estimate of the true niche breadth, since the true value will either be the same as or more extreme than the estimated value. However, to compare with other methods, we repeated calculations of niche breadth with precise plus imprecise records by matching imprecise records to: 1) the value at the geographic centroid of the area of likely collection; 2) the mean climatic value of the area of likely collection; and 3) the climatic value that is most different from the mean of the precise records. Using the climatic value farthest from the precise records, when contrasted with results using the nearest climatic value, allows us to examine the full breadth of climate space that the species could occupy. Matching records to the mean or centroid of the area of likely collection has been used in other works to accommodate use of imprecise records in species distribution modeling (Collins et al. 2017; Park & Davis 2017).

Climate change exposure

We used ecological niche models to estimate exposure to anticipated climate change. Full methods are described in Appendices 2 and 3 (the latter of which comprises a Overview, Data, Model, Assessment, and Prediction or ODMAP protocol; Zurell et al. 2020), so are only summarized here. Models were calibrated using the first 3 axes of the PCA on the 19 BIOCLIM variables. Background data were obtained from either an area delineated by a 300-km buffer around the minimum convex polygon of the occurrences used in each model, or from an area delineated using 300-km buffers around each individual occurrence. The former method assumes that a species can disperse to all intermediate locations between known occurrences, but

potentially overestimates the total area to which a species could disperse (e.g., when the range is disjunct). The latter method obviates issues with disjunct distributions, but by excluding environments that occur in gaps within a species' range, assumes that the species cannot occupy these environments or disperse to them (Barve et al. 2011). Both buffers were calculated separately for models using only precise records and for all records combined.

We modeled species' niches using Maxent ver. 3.3.3k (Phillips et al. 2006; Phillips & Dudík 2008) using precise and precise plus imprecise records. Models were then projected to present-day and future climate scenarios. For each species, we constructed 10 models (Fig. S2.1). For models based only on precise records, we calibrated one for each of the two definitions of background area. For models based on precise plus imprecise records, we calibrated eight models (2 background definitions × 4 methods for assigning climate variables to imprecise records). Each model was then projected to RCPs 4.5 and 8.5. We thresholded present-day and future predictions output using the value that excluded 10% of the occurrences from the set considered "present" (i.e., a 90% sensitivity rate). We then compared differences in area of several metrics indicative of area of habitat (Brooks et al. 2014) and exposure to climate change (Young et al. 2016) between model with and without imprecise records: 1) present-day area with suitable climate; 2) future suitable area; 3) stable area that is suitable present and in the future; and 4) loss and 5) gain in suitable climate area. Each of these metrics was calculated within a) the minimum convex polygon drawn around the precise records and b) within the minimum convex polygon drawn around all records, which was delineated as described above for calculation of EOO. Comparisons between models with and without imprecise records were conducted within each combination of background definition, emissions scenario, and method for assigning climate values to imprecise records. We compared differences in area of present-

day suitable climate, stable suitable climate, and loss and gain in climatically suitable area with paired Wilcoxon signed-rank tests.

Collection date of precise and imprecise records

Older records may be more imprecise so may tend to be discarded in favor of newer records with more precise locality information. To assess if imprecise records were generally older than precise records, we tested for differences between the median collection year of precise and imprecise records for each species using a Mann-Whitney U test.

Reproducibility

The analysis relied primarily on the sp (Bivand et al. 2013), rgeos (Bivand & Rundel 2020), geosphere (Hijmans 2019), dismo (Hijmans et al. 2017), raster (Hijmans 2021), and enmSdm (Smith 2020) packages for R Version 3.6.2 (R Core Team 2020). All scripts used in this analysis are available on the GitHub repository https://github.com/adamlilith/impreciseSpecimens_climateChangeVulnerability.

Results

Specimen data

The data downloaded from GBIF comprised 112,730 records (Table 1). Approximately one third of these records were missing coordinates entirely, and approximately another third were missing a value for coordinate uncertainty. Following application of our data cleaning procedures, removal of duplicate records, and elimination of species with fewer than 5 geographically unique precise records, we were left with 7.5% of the initial number of records, representing 32% of the

species (44 of 137) that occurred in the original data (Table 1). Records that could only be assigned to a geopolitical unit were the most abundant (Fig. 3a; median across retained species: 75 records, range: 6 to 519), followed by precise records (median: 25, range: 5 to 239), then records possessing coordinates with coordinate uncertainty >5000 m (median: 10, range: 0 to 101; Fig. 3 and Table S2.1).

Extent of occurrence

EOO calculated using precise plus imprecise was 86% larger (median across species) than when using precise records alone (range: 0 to 2011%; Fig. 4a; $P<10^{-12}$, Wilcoxon V=0). Including imprecise records at least doubled EOO for 34% of species (n=15 of 44), and at least tripled EOO for 27% of species (n=12). Species with fewer precise records tended to have the greatest increase in EOO when imprecise records were included (Fig. 4a).

Univariate and multivariate niche breadth

Including imprecise records increased estimated niche breadth for nearly all species, even when we conservatively used the climate value most similar to the mean across the precise records (Fig. 4b). Including imprecise records increased univariate niche breadth in MAT by a median value of 25% (range across species: 0 to 353%; $P<10^{-6}$, Wilcoxon V~0) and in MAP by 28% (range: 0 to 292%; $P<10^{-6}$; Wilcoxon V~0). Using other methods for assigning climate values to imprecise records increased niche breadths even more (Appendix S4). Species with the fewest precise records tended to have the greatest increase in univariate niche breadth when imprecise records were included (Fig. 4b).

Including imprecise records increased multivariate niche volume by a median value of 175% (range: 8 to 13,909%; Fig. 4c; $P<10^{-12}$, Wilcoxon V~0) and surface area of niche volumes by 79% (range: 3 to 1515%; $P<10^{-12}$, Wilcoxon V~0). Estimated niche breadths were larger for each of the other methods used to assign to assign climate values to imprecise records (Appendix S5). Species with the fewest records tended to have the greatest increase in niche volume and area when imprecise records were included (Fig. 4c).

Climate change exposure

We evaluated the effect of using different backgrounds for calibrating the niche models (buffered occurrences or buffers around minimum convex polygons surrounding occurrences), different climate scenarios (RCPs 4.5 and 8.5), and different areas in which climate change exposure was evaluated (within the convex polygon surrounding just precise occurrences or precise plus imprecise occurrences). Results were qualitatively similar, so in the main text we focus on models calibrated with background sites drawn from the buffered convex polygon projected to RCP4.5, with exposure calculated within the minimum convex polygon surrounding just precise records. Full results are presented in Appendices S6 and S7.

Including imprecise records increased estimated present-day climatically suitable area for nearly all species (Fig. 4a). Including imprecise records increased current climatically suitable area by 14% (median; range: -33 to 224%; Fig. 4a; $P<10^{-5}$, Wilcoxon V=113), with 90% of species experiencing an increase (39 of 44 species). For a few species climatically suitable area was lower when imprecise records were included because using imprecise records altered model parameterization and the threshold used to delineate suitable from unsuitable area, thereby changing the relative amount of area above and below the thresholds.

Including imprecise records increased the area predicted to remain climatically suitable (Fig. 5a) by 14% (median; range: -33 to 326%; $P<10^{-5}$, Wilcoxon V=120). The median percentage of current suitable area that was expected to be lost was similar (4% when using precise records versus 2% when using precise plus imprecise records; P=0.32, Wilcoxon V=531). However, models using only precise records predicted a greater range of loss than models using both kinds of records (Fig. 5b). In contrast, the area that is currently unsuitable but predicted to become suitable was larger when using just precise records (Fig. 5c; P=0.0129; Wilcoxon V=623).

Collection date of precise and imprecise records

We found no significant difference in the median age of the precise and imprecise occurrence records for 36 of the 44 species. For the other 8 species, the median age of the imprecise occurrences was more recent than the age of the precise occurrences (Mann-Whitney U <0.05; Appendix S8 Table S8.1).

Discussion

We found that including geospatially-imprecise increased estimates of EOO, climatic tolerances, and the area of suitable habitat in the present and the future (Figs. 4 and 5). Since we do not know the true locations of imprecise records, assigning a location or value of a climate variable will in most cases lead to an inaccurate match. Hence, our estimates of EOO, niche breadth, and exposure to climate change are admittedly inaccurate. Indeed, aversion to inaccuracy is the primary reason why imprecise records are commonly discarded before conducting conservation assessments (Fig. 1; Feeley & Silman 2010; Moudrý and Šímová 2012; Collins et al. 2017). However, our findings are based on using a conservative method for assigning locations and values of climate variables to imprecise records. Thus, it is reasonable to assume that our

estimates are still an underrepresentation of the true species geographic and environmental distributions. As a result, we contend that our results are a more accurate indicator of true EOO, climatic tolerances, and exposure to climate change because they are less of an underestimation of the true values of these metrics than metrics calculated using only precise records.

Implications for conservation

We found that even when using a conservative method for assigning imprecise records to spatial locations, EOO calculated using precise plus imprecise records was substantially larger (median increase across species: 86%) than when using precise records alone (Fig. 4a). If threats to species tend to act in a spatially autocorrelated manner (Fisher 2011), increasing the spatial spread of populations will reduce the probability that threats affect a large proportion of a species' populations (Reddinghaus & den Boer 1970; IUCN 2019). Hence, increasing the number of populations in an analysis of EOO will in general better reflect the true spatial spreading of risk among populations, even if some population cannot be located to precise locations (see Fig. 2a). Likewise, estimates of univariate niche breadth in MAT and MAP, and multivariate niche volume and area were also substantially larger when imprecise records were included (median increases ranging from 25 to 175%; Figs. 4b and c), indicating that many species may have wider climatic tolerances than would be apparent based on use of precise records alone. Similarly, niche models predicted that current climatically suitable area was larger when using precise plus imprecise records, even when the area of interest was restricted to the convex polygon surrounding just precise records (Fig. 4a; similar results were obtained using the polygon surrounding all records; Fig. S7.1). Areas that remain climatically suitable in the future were also larger (Fig. 5a and S7.2). In general, species with fewer precise records had greater

proportional increases in EOO, niche breadth and volume, and climatically suitable area in the present and future.

Indeed, ignoring imprecise records can misconstrue outcomes of conservation vulnerability assessments. For example, under IUCN Red List criterion B1, species qualify as threatened if they have an EOO <100 km^2 (in addition to other criteria; IUCN 2019; see also Young et al. 2016). While none of the species in our analysis had an extent of occurrence <100 km^2 when using just precise records, it is certainly possible that this threshold could be crossed in assessments of other species. Indeed, this may be the case for other species of *Asclepias* in North America, as we restricted our analysis to the 44 species with ≥5 geographically unique precise records out of a possible 137 in the original data (only a few of which are non-native). Since precise records typically encompass a smaller region than the combination of precise and imprecise records (Fig. 4a), it is reasonable to assume that excluding imprecise records could increase the chance that a species would have an EOO smaller than this threshold, thus making it appear to meet this criterion.

Our intent here is to provoke a reconsideration of the benefits and costs of discarding spatially imprecise records when assessing species' conservation status (Table 2). The trade-offs between these choices are based on the philosophical approach under which any particular conservation assessment is conducted. First, when decisions must be made that affect whether or not a species is designated as vulnerable, many assessments adopt a precautionary strategy that errs on the side of assuming a species is more vulnerable (Moyle 2005). Indeed, a precautionary approach is generally advised when assessing current (IUCN 2019) and potential future (Huntley et al. 2016a) vulnerability. Since discarding spatially imprecise records reduces apparent niche breadth and extent of distribution (Fig. 4), ignoring imprecise records inherently aligns with a

precautionary approach to assessment. The advantage of a precautionary approach is that it ensures few truly vulnerable species are mistakenly classified as “not vulnerable” because species tend to be assigned a vulnerability status that is at least as severe as their true status. The opposite of a precautionary approach is an evidentiary approach, which aims to classify species as vulnerable only if there is strong evidence to support such a designation (IUCN 2019). The evidentiary approach would thus seem to align with inclusion of imprecise records because their use broadens species’ apparent geographic and environmental distributions. As a result, ignoring imprecise records would generally seem advisable.

We contend that spatially imprecise records—despite their limitations—can represent valid information on species’ distributions and environmental tolerances. As such, unless there is sound reason for doing so, ignoring removing imprecise records seems arbitrary and capricious. Current investment in conservation is inadequate for protecting biodiversity against present threats (Waldron et al. 2017). As a result, targeting species that are indeed threatened is critical for ensuring conservation effort optimizes return on investment (Murdoch et al. 2007). Hence, overstating vulnerability, and thus ignoring imprecise records, is counterproductive to the overall goal of protecting life on Earth. We thus argue for a re-assessment of the default practice of removing imprecise records before conducting conservation assessments. We realize that this position seems evidently in approach. However, evidentiary and precautionary approaches are not inherently antagonistic across all levels of analysis. For example, our methods for assigning locations and environmental values to imprecise records are inherently precautionary in that the species’ true distributions and tolerances will most likely be broader than our estimates.

Regardless of whether a precautionary or evidentiary approach is adopted, conservation assessments are most informative when they account for all relevant aspects of uncertainty

(Reece et al. 2013; Huntley et al. 2016b; IUCN 2019). The decisions over whether to retain or discard imprecise records, and indeed, over what constitutes an "imprecise" record, represent key aspects of uncertainty. Nevertheless, across all of the articles that we reviewed which reported discarding records based on coordinate uncertainty, none of them evaluated the consequences of this choice (Fig. 1). Decisions over how to delineate precise from imprecise records, and whether or not to retain the latter, are often determined by the resolution of available environmental data (Graham et al. 2008). However, these decisions are not necessarily replicable in the sense that another set of assessors may make an equally defensible, yet different decision. Ignoring the subjectivity inherent in these decisions reduces apparent variation in the final assessment of species' distributions and environmental tolerances, and thus does not provide a full account of underlying uncertainty (Huntley et al. 2016b). Acknowledging uncertainty is important even in cases where inclusion or exclusion of imprecise records would not change a species' overall qualitative vulnerability status. For example, when allocating scarce conservation resources to two species with the same qualitative conservation status, the species with greater uncertainty in its status will present different trade-offs compared to the one with less uncertainty (Smith et al. 2016). As a result, we advise accounting for uncertainty inherent in the decision to include versus exclude of imprecise occurrences from conservation vulnerability assessments, or at least to fully consider trade-offs when adopting one strategy over the other (Table 2).

Methodological considerations

The manner in which imprecise records can be accommodated depends on the approach used to translate occurrences to climatic tolerances. The simplest method entails calculating niche breadth using the range of extreme environments occupied by a species (e.g., by overlaying range maps onto climate surfaces; Foden et al. 2013), similar to the way we calculated univariate

niche breadth in MAT and MAP (Figs. 2b and 4b). Assessors can bracket possible outcomes by calculating climatic tolerances based on just precise occurrences and on precise plus imprecise coordinates. Of course, this requires a method for matching imprecise occurrences to environmental data (e.g., means, centroids, or most or least similar values within each polygon encompassing an imprecise occurrence; Park & Davis 2017), which is another source of uncertainty that should be acknowledged. Alternatively, assessors can employ modeling methods that accommodate imprecise occurrences (Hefley et al. 2014 and 2017; Velásquez-Tibatá et al. 2016) or spatial tools that assess the potential effect of coordinate uncertainty on model outcome (Naimi et al. 2014). Some of these methods allow uncertainty in location to propagate through to the final coefficient estimates, thereby implicitly accounting for positional error. While such methods are potentially the best means of incorporating imprecise records, their use remains limited to date (we found no examples using these methods in our sampling of the literature; Appendix S1).

The effect of imprecise records on niche breadth depends on how well precise records sample geographic and niche space (Moudrý & Šímová 2012). Generally, species that have few records and/or narrower niche breadth will be affected more by the inclusion of imprecise records (Tulowiecki et al. 2015; Collins et al. 2017; Velásquez-Tibatá et al. 2016; but see Gábor et al. 2020). This was indeed true in our case. For example, we found that including imprecise records increased EOO by more than five-fold for some species with <80 records (Fig. 4a). The degree of spatial autocorrelation in environmental conditions can ameliorate the effects of positional uncertainty if high autocorrelation causes mis-located records to be associated with environmental values similar to the values at their true location of collecting (Naimi et al. 2014), so long as the errors themselves are not environmentally structured (Hefley et al. 2017). Even

when spatial autocorrelation is low (i.e., large environmental heterogeneity), using the nearest climate datum as we did here minimizes risks of over-inflating climatic tolerances.

Owing to advances in collection practices, mapping, and GPS (Wieczorek et al. 2004; Wieczorek & Chapman 2020), imprecise records may tend to be older than precise records. If this is the case, two opposing considerations should also inform the decision over whether to include imprecise records. First, matching older specimens to average climate across a period may misrepresent the actual mean climate experienced by the species prior to collection (Garcia et al. 2019). Given that change in climate is accelerating (Smith et al. 2015), older specimens represent populations that have since been extirpated due to climate change that has occurred in the latter portion of a period over which climate has been averaged. If this is the case, including older records may overstate environmental tolerances (cf. Escribano et al. 2016). In contrast, older specimens may represent populations that have since been extirpated due to non-climatic threats even though their site of collection may still remain climatically suitable. As a result, excluding them would underestimate climatic tolerances (Faurby & Araujo 2018). Both of these possibilities relate to the previous discussion about a precautionary versus evidentiary approach and about accounting for uncertainty, so decisions should be made in that regard. In our analysis, we did not find that imprecise specimens were older than precise specimens (Appendix S8), but this may not always be the case.

Reproducibility in data processing and cleaning

Regardless of the decisions researchers make about which records to include in their analyses, it is important to clearly document the criteria and steps used to clean specimen data prior to conducting environmental niche modelling (Chapman 2005; Wieczorek & Chapman 2020).

Remarkably, only about half of the papers that we reviewed described their data cleaning process. This is troubling because it means either that data were used without proper screening for issues common to biodiversity databases (Feeley & Silman 2010; Gueta & Carmel 2016; Meyer et al. 2016; Bloom et al. 2018; Moudrý & Devillers 2020), or that the analyses were not fully described. Given the reproducibility crisis in science in general and ecology specifically (Powers & Hampton 2019; Culina et al. 2020), we strongly advocate detailed descriptions of data cleaning procedures, which could serve as extensions to recently-published metadata standards for niche modeling (e.g., Feng et al. 2019; Zurell et al. 2020). We also note that even when scripts and input/output data are provided by authors, comments in computer code usually do not explain why particular filters were chosen, so we advise accompanying such materials with textual descriptions of these rationale.

Conclusions

To date, the general consensus has been to discard imprecise records because their use can inflate estimates of niche breadth and reduce the apparent accuracy of ecological niche models. This widespread aversion to using spatially imprecise records may be due in part to the fact that nearly all studies evaluating the effects of coordinate imprecision do so by adding spatial error to erstwhile precise records (e.g., Graham et al. 2008; Fernandez et al. 2009; Osborne & Leitão 2009; Gueta & Carmel 2016; Mitchell et al. 2016; Hefley et al. 2017; Soultan & Safi 2017; Tulowiecki et al. 2015; Gábor et al. 2020). Although this approach keeps sample sizes constant between treatments with or without spatially imprecise records, it is not reflective of real-world situations where assessors usually start with a mix of relatively precisely- and imprecisely-geolocated records but must decide how to delineate the two groups and whether or not to discard the imprecise ones.

We advocate for a re-consideration over whether spatially imprecise occurrence records should always be excluded from conservation assessments. Using only precise records tends to result in smaller estimated extents of occurrence and niche breadth, as well as greater vulnerability to climate change. The decision over how to define imprecise records and whether or not to use them is an important component of the overall uncertainty in inherent in the outcome of any conservation assessment relying on museum or herbarium specimen data. Discarding imprecise records ignores this critical aspect of uncertainty.

We have contextualized our results in light of assessing species' conservation status, but specimen records are used to answer many questions in ecology and evolution, including investigations of range size, biogeographic history, niche conservatism, and speciation (e.g. Fisher-Reid et al. 2012; Quintero & Wiens 2013; Slatyer et al. 2013; Hoban et al. 2019). Thus, nearly any analysis relying on specimen data to ascertain range extent or niche breadths will benefit from a full consideration of the benefits and costs of using spatially imprecise records, and from a full accounting of the uncertainty inherent in decisions over whether and how to include or exclude imprecise records.

Acknowledgements

DH was supported by the Department of Biology of Washington University in Saint Louis. This project was made possible in part by the Institute of Museum and Library Services (FAIN MG-30-15-0094-15) to ABS and by the Alan Graham Fund in Global Change.

Conflict of Interest Statement

The authors declare no conflict of interest.

Biosketch

The Global Change Conservation Laboratory at the Missouri Botanical Garden identifies solutions to pressing environmental problems by leveraging precise and the imprecise information on biodiversity. We are inspired by the cumulative person-millennia of field work and curatorial attention devoted to amassing collection records of Earth's species and the exigency of using this data to its fullest potential.

Literature cited

Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S.P. Maher, A.T. Peterson, J. Soberón, and F. Villalobos. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecological Modeling 222:1810-1819.

Bivand, R. and Rundel, C. 2020. rgeos: Interface to Geometry Engine: Open Source ('GEOS'). Version 0.5-5. https://CRAN.R-project.org/package=rgeos

Bivand, R., Pebesma, E., and Gómez-Rubio, V. 2013. Applied Spatial Data Analysis with R. Springer-Verlag, New York.

Bloom, T.D.S., Flower, A., and DeChaine, E.G. 2018. Why georeferencing matters: Introducing a practical protocol to prepare species occurrence records for spatial analysis. Ecology and Evolution 8:765-777.

Brooks, T.M., Pimm, S.L., Akçakaya, R., Buchanan, G.M., Butchart, S.H.M., Foden, W., Hilton-Taylor, C., Hoffmann, M., Jenkins, C.N., Joppa, L., Li, B.L., Ocampo-Peñuela, N., and Rondinini, C. 2014 Measuring terrestrial area of habitat (AOH) and its utility for the IUCN Red List. Trends in Ecology and Evolution 34:977-986.

Chapman A.D. and Wieczorek, J.R. 2020 Georeferencing Best Practices. Copenhagen: GBIF Secretariat. https://doi.org/10.15468/doc-gg7h-s853Chapman 2005.

Cheng, Y., Tjaden, N.B., Jaeschke, A., Thomas, S.M, and Beierkuhnlein, C. 2021. Using centroids of spatial units in ecological niche modeling: Effects on model performance in the context of environmental data grain size. Global Ecology and Biogeography 30:611-621.

Collins, S.D., Abbott, J.C., and McIntyre, N.E. 2017. Quantifying the degree of bias from using county-scale data in species distribution modeling: Can increasing sample size or using county-averaged environmental data reduce distributional overprediction? Ecology and Evolution 7:6012-6022.

Culina, A., Baglioni, M., Crowther, T.W., Visser, M.E., Woutersen-Windhouwer, S., and Manghi, P. 2019. Navigating the unfolding open data landscape in ecology and evolution. Nature Ecology and Evolution 2:420-426.

Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C., and Mace, G.M. 2011. Beyond predictions: Biodiversity conservation in a changing climate. Science 332:53-58.

Escribano, N., Ariño, A.H., and Galacia, D. 2016. Biodiversity data obsolescence and land use changes. PeerJ 4:e2743.

Faurby, S. and Araújo, M.B. 2018. Anthropogenic range contractions bias species climate change forecasts. Nature Climate Change 8:252-256.

Feeley, K.J. and Silman, M.R. 2010. Modelling the responses of Andean and Amazonian plant species to climate change: The effects of georeferencing errors and the importance of data filtering. Journal of Biogeography 37:733-740.

Feeley, K.J. and Silman, M.R. 2011. Keep collecting: Accurate species distribution modeling requires more collections than previously thought. Diversity and Distributions 17:1132-1140.

Feng, X., Park, D.S., Walker, C., Peterson, A.T., Merow, C., and Papeş, M. 2019. A checklist for maximizing reproducibility of ecological niche models. Nature Ecology and Evolution 3:1382-1395.

Fernandez, M.A., S.D. Blum, S. Reichle, Q. Guo, B. Holzman, and H. Hamilton. 2009. Locality uncertainty and the differential performance of four common niche-based modeling techniques. Biodiversity Informatics 6:36-52.

Fick, S.E. and Hijmans, R.J. 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37:4302-4315.

Fisher, D.O. 2011. Trajectories from extinction: Where are missing mammals rediscovered? Global Ecology and Biogeography 20:415-425.

Fisher-Reid, M.C., Kozak, K.H., and Wiens, J.J. 2012. How is the rate of climatic-niche evolution related to climatic-niche breadth? Evolution 66:3836-3851.

Foden, W.B. and Young, B.E. (eds.) 2016. IUCN SSC Guidelines for Assessing Species' Vulnerability to Climate Change. Version 1.0. Occasional Paper of the IUCN Species Survival Commission No. 59. Cambridge, UK and Gland, Switzerland: IUCN Species Survival Commission. x+114 pp.

Foden, W.B., Mace, G.M., and Butchart, S.H.M. 2013. Indicators of climate change impacts on biodiversity. Pp. 120-137 in Collin, B., Pettorelli, N., Baillie, J.E.M., and Durant, S.M. (eds.) Biodiversity Monitoring and Conservation: Bridging the Gap between Global Commitment and Local Action, 1st ed. John Wiley and Sons, Indianapolis.

Gábor, L., Moudrý, V., Lecours, V., Malavasi, M., Barták, V., Fogl, M., Šímová, P., Rocchini, D., and Václavík, T. 2020. The effect of positional error on fine scale species distribution models increases for specialist species. Ecography 43:256-269.

Garcia, R.A., Allen, J.L., and Clusella-Trullas, S. 2019. Rethinking the scale and formulation of indices assessing organism vulnerability to warmer habitats. Ecography 42:1024-1036.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A., and the NCEAS Predicting Species Distributions Working Group. 2008. The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology 45:239-247.

Gueta, T. and Carmel, Y. 2016. Quantifying the value of user-level cleaning for big data: A case study using mammal distribution models. Ecological Informatics 34:139-145.

Hefley, T.J., Baasch, D.M., Tyre, A.J., and Blankenship, E.E. 2014. Correction of location errors for presence-only species distribution models. Methods in Ecology and Evolution 5:207-214.

Hefley, T.J., Brost, B.M., and Hooten, M.B. 2017. Bias correction of bounded location errors in presence-only data. Methods in Ecology and Evolution 8:1566-1573.

Hijmans, R.J. 2019. geosphere: Spherical Trigonometry. R package version 1.5-10. https://CRAN.R-project.org/package=geosphere.

Hijmans, R.J. 2016. raster: Geographic Data Analysis and Modeling. R package version 2.5-8. https://CRAN.R-project.org/package=raster.

Hijmans, R.J., Phillips, S.J., Leathwick, J., and Elith, J. 2020. dismo: Species Distribution Modeling. R package version 1.3-3. https://CRAN.R-project.org/package=dismo.

Hoban, S., Dawson, A. Robinson, J., Smith, A.B., Strand, A. 2019. Inference of biogeographic history by formally integrating distinct lines of evidence: genetic, environmental niche, and fossil. Ecography 42:1991-2011.

Huntley, B., Foden, W.B., Smith, A.B., Platts, P., Watson, J. and Garcia, R.A. 2016. Chapter 5. Using CCVAs and interpreting their results. In W.B. Foden and B.E. Young, editors. IUCN SSC Guidelines for Assessing Species' Vulnerability to Climate Change. Version 1.0. Occasional Paper of the IUCN Species Survival Commission No. 59. Gland, Switzerland and Cambridge, UK. pp 33-48.

Huntley, B., Foden, W.B., Pearce-Higgins, J., and Smith, A.B. 2016. Chapter 6. Understanding and working with uncertainty. In W.B. Foden and B.E. Young, editors. IUCN SSC Guidelines for Assessing Species' Vulnerability to Climate Change. Version 1.0. Occasional Paper of the IUCN Species Survival Commission No. 59. Gland, Switzerland and Cambridge, UK. pp 49-56.

IUCN Standards and Petitions Committee. 2019. Guidelines for Using the IUCN Red List Categories and Criteria. Version 14. Prepared by the Standards and Petitions Committee. Downloadable from http://www.iucnredlist.org/documents/RedListGuidelines.pdf (2021-05-24).

Lomba, A., L. Pellissier, C. Randin, J. Vicente, J. Horondo, and A. Guisan. 2010. Overcoming the rare species modeling complex: A novel hierarchical framework applied to an Iberian endemic plant. Biological Conservation 143:2647-2657.

Marcer, A., Haston, E., Groom, Q., et al. 2021 Quality issues in georeferencing: From physical collections to digital data repositories for ecological research. Diversity & Distributions 27:564-567.

Meyer, C., Wiegelt, P., and Kreft, H. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters 19:992-1006.

Miller, J.S., Porter-Morgan, H.A., Stevens, H., Boom, B., Krupnick, G.A., Acevedo-Rodríguez, P., Fleming, J., and Gensler, M. 2012. Addressing targets two of the Global Strategy for Plant Conservation by rapidly identifying plants at risk. Biodiversity Conservation 21:1877-1887.

Mitchell, P.J., Monk, J., and Laurenson, L. 2016. Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across sample sizes. Methods in Ecology and Evolution 8:12-21.

Moudrý, V. and Šímová, P. 2012. Influence of positional accuracy, sample size and scale on modeling species distributions: A review. International Journal of Geographic Information Science 26:2083-2095.

Moudrý, V. and Devillers, R. 2020. Quality and usability challenges of global marine biodiversity databases: An example for marine mammal data. Ecological Informatics 56:101051.

Moyle, B. 2005. Making the Precautionary Principle work for biodiversity: Avoiding perverse outcomes in decision-making under uncertainty. Pp. 159-172 in Cooney, R. and B. Dickson (eds.) Biodiversity and the Precautionary Principle: Risk and Uncertainty in Conservation and Sustainable Use. Earthscan, London. 314 pp.

Murdoch, W., S. Polasky, K.A. Wilson, H. Possingham, P. Kareiva, and R. Shaw. 2007. Maximizing return on investment in conservation. Biological Conservation 139:375-388.

Naimi, B., Hamm, N.A.S., Groen, T.A., Skidmore, A.K., and Toxopeus, A.G. 2014. Where is positional uncertainty a problem for species distribution modeling? Ecography 57:191-203.

Nix, H.A. 1986. A biogeographic analysis of Australian elapid snakes. Atlas of elapid snakes of Australia: Australian flora and fauna series 7 (ed. by R. Longmore), pp. 4-15. Bureau of Flora and Fauna, Canberra.

O'Neill, B.C., Kreigler, E., Ebi, K.L., Kemp-Benedict, E., Riahi, K., Rothman, D.S., van Ruijven, B., van Vuuren, D.P., Birkman, J., Kok, K., Levy, M., and Solecki, W. 2017. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. Global Environmental Change 42:169-180.

Osborne, P.E. and Leitão, P.J. 2009. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. Diversity and Distributions 15:671-681.

Pacifici, M., Foden, W.B., Visconti, P., Watson, J.E.M., Butchart, S.H.M., Kovacs, K.M., Scheffers, B.R., Hole, D.G., Martin, T.G., Akçakaya, H.R., Corlett, R.T., Huntley, B., Brickford, D., Carr, J.A., Hoffmann, A.A., Midgley, G.F., Pearce-Kelly, P. Pearson, R.G., Williams, S.E., Willis, S.G., Yoing, B., and Rondinini, C. 2015. Assessing species vulnerability to climate change. Nature Climate Change 5:215-225.

Park, D.S. and Davis, C.C. 2017. Implications and alternatives of assigning climate data to geographical centroids. Journal of Biogeography 44:2188-2198.

Pender, J.E., Hipp, A.L., Hahn, M., Kartesz, J., Nishino, M., and Starr, J.R. 2019. How sensitive are climatic niche inferences to distribution data sampling? A comparison of Biota of North America Program (BONAP) and Global Biodiversity Information Facility (GBIF) datasets. Ecological Informatics 54:100991.

Peterson, A.T., Cobos, M.E., and Jiménez-García, D. 2018. Major challenges for correlational ecological niche model projections to future climate conditions. Annals of the New York Academy of Sciences 1429:66-77.

Phillips, S.J. and Dudík, M. 2008. Modeling species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography 31:161-175.

Phillips, S.J., Anderson, R.P., and Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.

Powers, S.M. and Hampton, S.E. 2019. Open science, reproducibility, and transparency in ecology. Ecological Applications 29:e01822.

Qiao, H., Feng, X., Escobar, L.E., Peterson, A.T., Soberón, J., Zhu, G., and Papeş. 2019. An evaluation of transferability of ecological niche models. Ecography 42:521-534.

Quintero, I. and Wiens, J.J. 2013. What determines the climatic niche width of a species? The role of spatial and temporal climatic variation in three vertebrate clades. Global Ecology and Biogeography 22:422-432.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reddinghaus, J. and P.J. den Boer. 1970. Simulation experiments illustrating stabilization of animal numbers by spreading of risk. Oecologia 5:240-284.

Reece, J.S., Noss, R.F., Oetting, J., Hoctor, T., and Volk, M. 2013. A vulnerability assessment of 300 species in Florida: Threats from sea level rise, land use, and climate change. Public Library of Science ONE 8:e80658.

Sheth, S.N., Lohmann, L.G., Distler, T., and Jiménez, I.. 2012. Understanding bias in geographic range size estimates. Global Ecology and Biogeography 21:732-742.

Slatyer, R.A., Hirst, M., and Sexton, J.P. 2013. Niche breadth predicts geographical range size: A general ecological pattern. Ecology Letters 16:1104-1114.

Smith, A.B. 2020. enmSdm: Tools for modeling niches and distributions of species. R package version 0.5.3.5. http://github.com/adamlilith/enmSdm

Smith, A.B., Long, Q.G., and Albrecht, M.A. 2016. Shifting targets: spatial priorities for ex situ plant conservation depend on interactions between current threats, climate change, and uncertainty. Biodiversity and Conservation 25:905-922.

Smith, S.J., Edmonds, J., Hartin, C.A., Mundra, A., and Calvin, K. 2015. Near-term acceleration in the rate of temperature change. Nature Climate Change 5:333-336.

Soultan, A. and Safi, K. 2017. The interplay of various sources of noise and reliability on species distribution models hinges on ecological specialization. Public Library of Science ONE 12:e0187906.

Thuiller, W., Araújo, M.B., Pearson, R.G., Whittaker, R.J., Brotons, L., and Lavorel, S. 2004. Uncertainty in predictions of extinction risk. Nature 430:33.

Tulowiecki, S.J., Larsen, C.P.S., and Wang, Y-C. 2015. Effects of positional error on modeling species distributions: A perspective using presettlement land survey records. Plant Ecology 216:67-85.

Velásquez-Tibatá, J., Graham, C.G., and Munch, S.B. 2016. Using measurement error models to account for georeferencing error in species distribution models. Ecography 39:305-316.

Waldron, A., Miller, D.C., Redding, D., Mooers, A., Kuhn, T.S., Nibbelink, N., Roberts, J.T., Tobias, J.A., and Gittleman, J.L. 2017. Reductions in global biodiversity loss predicted from conservation spending. Nature 551:364-367.

Wieczorek, J., Guo, Q., and Hijmans, R.J. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. International Journal for Geographical Information Science 18:745-767.

Williams, S.E., Shoo, L.P., Isaac, J.L., Hoffmann, A.A., and Langham, G. 2008. Towards an integrated framework for assessing the vulnerability of species to climate change. Public Library of Science Biology 6:e325.

Young, B.E., Byers, E., Hammerson, G., Frances, A., Oliver, L., and Treher, A. 2016. Guidelines for Using the NatureServe Climate Change Vulnerability Index, Version 3.02. NatureServe, Arlington.

Zizka, A., ter Steege, H., Pessoa, M. de C.R, and Antonello, A. 2018. Finding needles in the haystack: Where to look for rare species in the American tropics. Ecography 41:321-330.

Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitao, P., Park, D.S., Peterson, A.T., Rapacciulo, G., Schmatz, D.R., Schroder, B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., and Merow, C. 2020. A standard protocol for reporting species distribution models. Ecography 43:1261-1277.

Table 1. Number and percentage of records of North American *Asclepias* in GBIF as of November, 2019. There were 112,730 records in the raw data. Just 7.5% of these records were usable for this analysis. * Larger than the area of San Bernardino county, California (52,104.5 km^2).

	Records	
Issue/status	**Number**	**Percent of total**
Problems with data (categories are not mutually exclusive)		
Invalid species name	987	0.9%
Invalid collection year	9,833	8.7%
Missing coordinates	35,743	31.7%
Possess coordinates but missing coordinate uncertainty	36,123	32.0%
Missing state/province	1,835	1.6%
Missing county	57,537	51.0%
Administrative mismatch (e.g., coordinates not in county)	3,676	3.3%
No information in locality fields	12,121	10.8%
Cultivated/purchased/not naturally occurring	586	0.5%
Study-specific issues (categories are not mutually exclusive)		
Observational record	45,939	40.8%
Collected before 1970	29,476	26.1%
Geographic duplicates	14,099	12.5%
Area of uncertainty too large*	7,688	6.8%
Outside coterminous North America and minor outlying islands	89	0.1%
Classifications across all records and species (mutually exclusive categories)		
Precise	34,327	30.5%
Imprecise, with coordinate uncertainty >5000 m	5,993	5.3%
Imprecise, locatable to county level	58,583	52.0%
Imprecise, locatable to state level	9,927	8.8%
Unclassifiable	3,900	3.5%
Total records	112,730	
Non-duplicate, usable records for species with ≥5 precise records (mutually exclusive)		
Precise	2,139	1.9%
Imprecise, with coordinate uncertainty >5000 m	889	0.8%
Imprecise, locatable to geopolitical unit ≤ critical size*	5,452	4.8%
Total usable	8,480	7.5%

Table 2. Trade-offs between using versus discarding spatially imprecise occurrence records when assessing conservation status from range size (e.g., extent of occurrence), niche breadth, and environmental tolerances from occurrences.

	Using precise only	Using precise plus imprecise
Sample size	Reduced	Increased
Matching to values of environmental variables	With little error	Indeterminate, possibly with large error
Sampling of geographic and niche space	Restricted	More complete
Method of analysis	Simpler	More complicated
Estimated niche breadth	Potentially larger underestimation	Potentially smaller underestimation
Climate change exposure	Species appear more vulnerable	Species appear less vulnerable
Approach	Precautionary (fewer regrets because more species receive attention)	Evidentiary (better return on conservation investment because only truly vulnerable species receive attention)

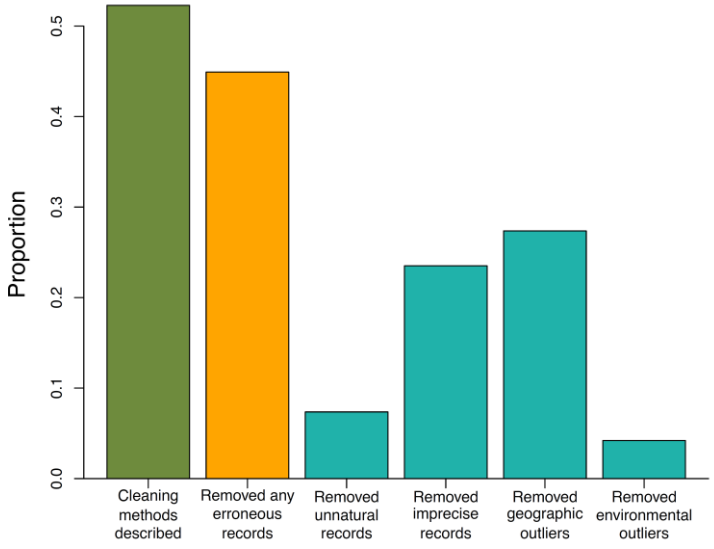

Figure 1. To characterize the methods used to address uncertainty in imprecisely-geolocated occurrences, we reviewed 1,950 peer-reviewed articles published between 2010 and 2019 that used specimen records and ecological niche modeling (full methods described in Appendix S1). Of 285 relevant articles, only 52% described any methods for cleaning data. Across all publications, 45% addressed issues related to coordinate uncertainty by removing records before the analysis. Bases for removing records included being "unnatural" (e.g., cultivated or purchased), having imprecise or unknown coordinate uncertainty, lying outside the species' presumed range, or falling outside the environmental tolerances of the species. No studies we reviewed used modeling methods that could account for uncertainty in spatial location or explicitly indicated that coarser resolution environmental data was used to accommodate spatially imprecise records.

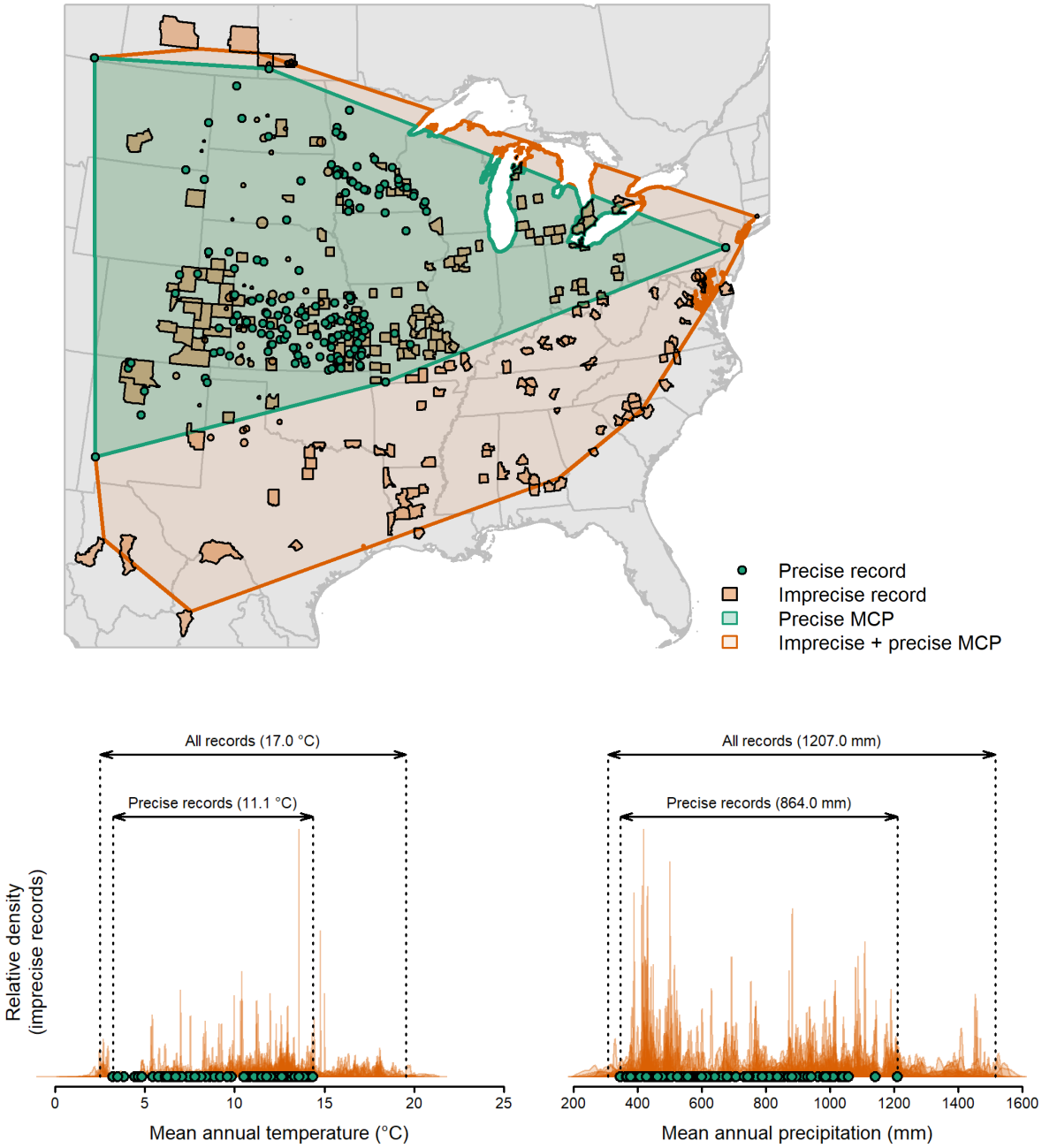

Figure 2. An example of differences in extent of occurrence (EOO) and univariate niche breadth measured using precise versus precise plus imprecise records of *Asclepias viridiflora*. a) EOO estimated using either just precise or precise plus imprecise records. Imprecise records are represented by the smallest geopolitical unit to which a record can be located (a county or a state), or by circles represented by buffering a coordinate by coordinate uncertainty. EOO is

calculated from the minimum convex polygon around just precise records, or around precise records plus (for imprecise records) the point in the record's polygon closest to the centroid of the precise records. Projection: Albers conic, equal-area. b) and c) Difference in climatic niche breadth of *A. viridiflora* for mean annual temperature and precipitation estimated using precise versus precise plus imprecise records. The value of temperature or precipitation at each precise record is represented by a point (green) and the distribution of temperature or precipitation across all locations encompassed by each imprecise record is represented by a smoothed density kernel, one per record (orange). For precise records, climatic niche breadth is defined as the range of temperature or precipitation encompassed by the records. For precise plus imprecise records, niche breadth is the range encompassed by the precise records and the values in each imprecise record's distribution that are closest to the mean value of the precise records.

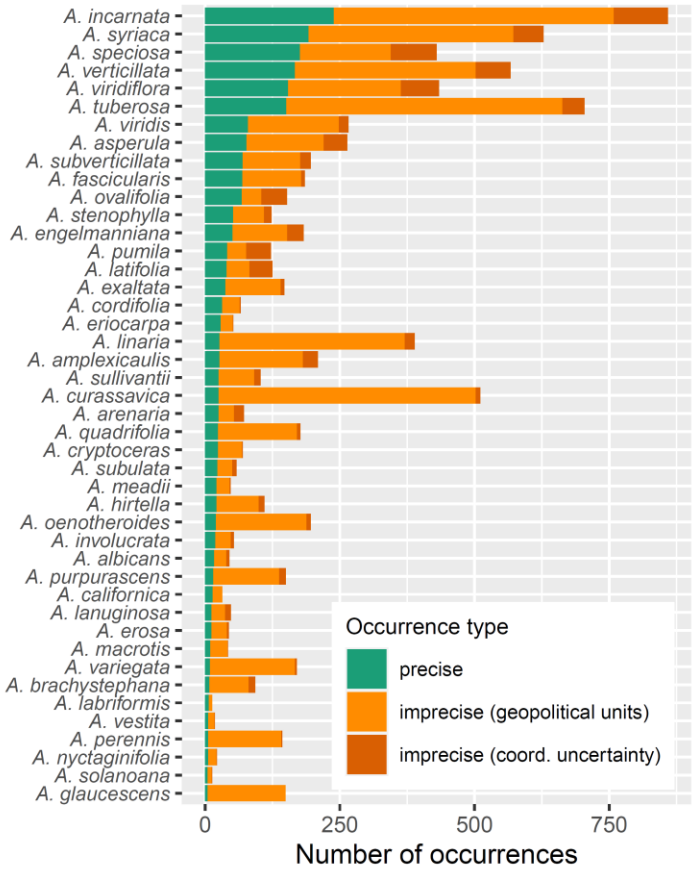

Figure 3. Number of non-duplicate records by type for each species after data cleaning. Imprecise records are defined as records with a coordinate uncertainty >5000 m, or records that can only be located with confidence to a geopolitical unit.

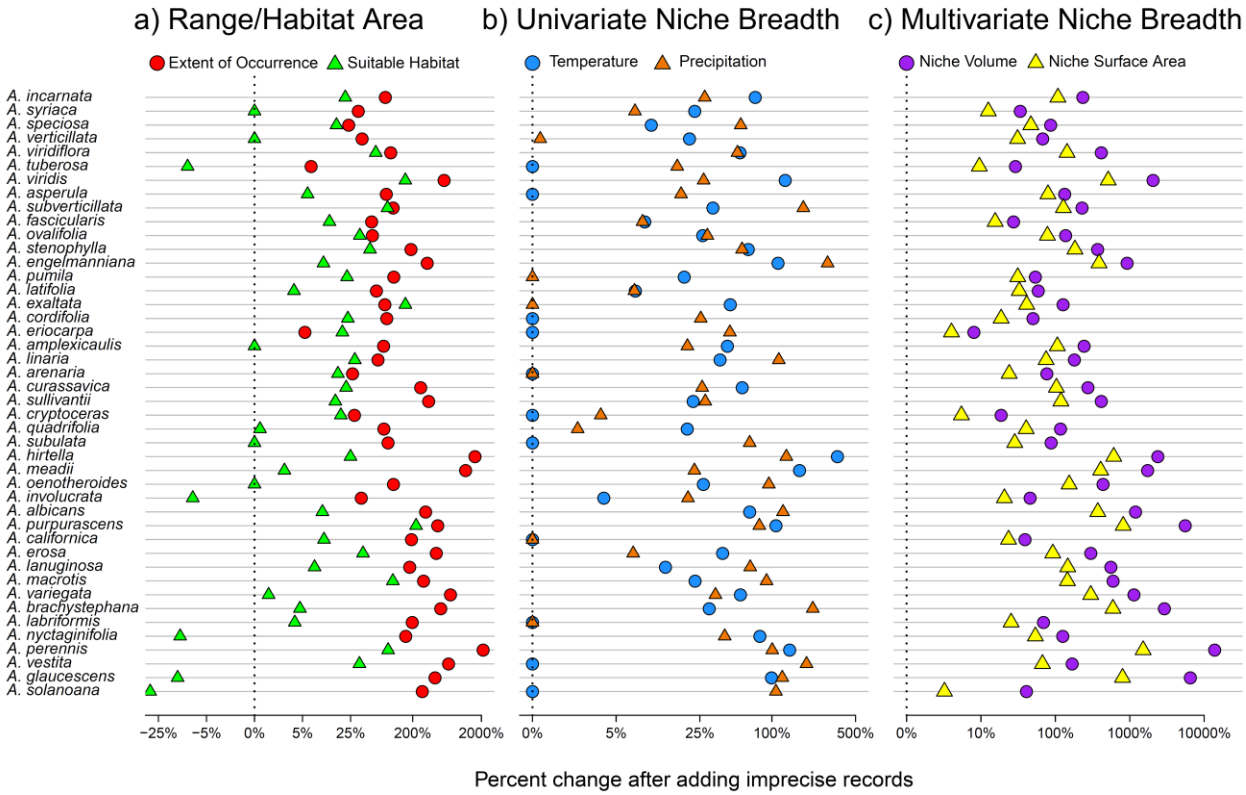

Figure 4. Including imprecise records increases estimated geographic range size and niche breadth. a) Change in extent of occurrence and present-day climatically suitable habitat estimated using Maxent. b) Change in univariate niche breadth in mean annual temperature and mean annual precipitation. c) Change in multivariate niche volume and surface area of this volume. Species are sorted from top to bottom from most to least number of precise records. Note the log scale along the x-axis.

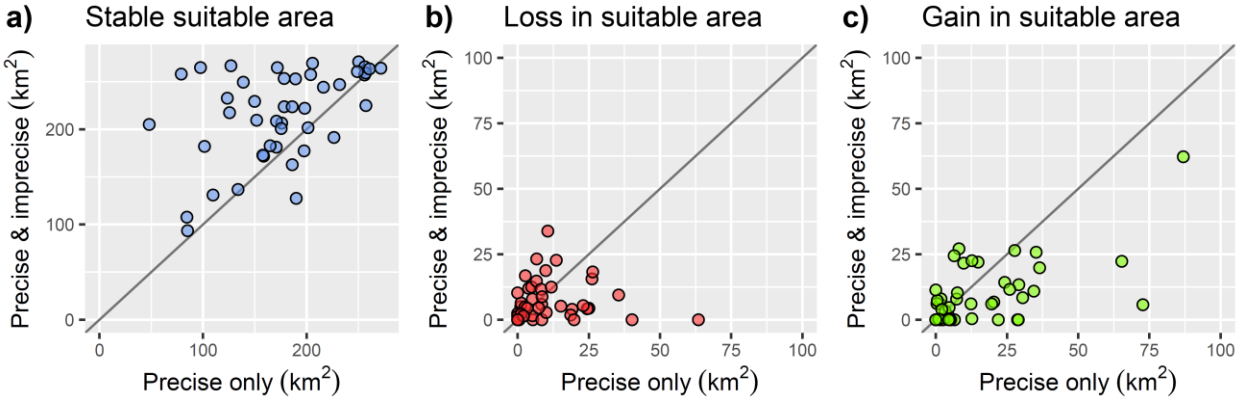

Figure 5. Differences in stable climatically suitable area, and in loss and gain in suitable area between ecological niche models using just precise records versus precise plus imprecise records. a) Area predicted to be suitable in the present and future. b) Climatically suitable area predicted to be lost due to climate change. c) Area predicted to be gained.